

COMPARAISON DE TROIS METHODES D'ANALYSE DE VARIABLES LATENTES LONGITUDINALES

Myriam Blanchin, Jean-Benoit Hardouin, Tanguy Le Neel, Gildas Kubis, Véronique Sébille

EA 4275 "Biostatistique, Recherche clinique et Mesures Subjectives en Santé"
UFR de Pharmacie, Université de Nantes

Résumé :

L'évaluation de mesures subjectives, comme la qualité de vie, prend une place de plus en plus importante en recherche clinique et en épidémiologie. Dans ce domaine, les données sont fréquemment recueillies de manière longitudinale afin de permettre de décrire l'évolution du critère d'intérêt dans le temps. La méthode la plus adéquate pour analyser des mesures subjectives recueillies longitudinalement reste à identifier. Elle peut être basée sur l'approche classique ou l'approche Théorie de Réponse aux Items (IRT) et doit prendre en compte le caractère latent du critère d'intérêt ainsi que la spécificité des données longitudinales.

Une étude de simulation a été mise en place afin de comparer trois méthodes d'analyse de variables latentes longitudinales. Deux de ces méthodes sont basées sur la variable latente (IRT). La dernière méthode repose sur le score (approche classique).

Trois critères de performance ont été évalués afin de pouvoir comparer chaque méthode: le risque de première espèce, la puissance et le biais des estimations. Ces critères peuvent être affectés par : (i) l'effectif de l'échantillon, (ii) le nombre d'items, (iii) la corrélation entre les deux mesures et (iv) la variance de la variable latente. Ainsi, plusieurs scénarios ont été envisagés combinant différentes valeurs de paramètres.

Cette étude montre des résultats différents en termes de risque de première espèce et de puissance pour les trois méthodes car elles sont toutes affectées par la valeur des quatre paramètres déjà cités.

Mots-clés : modèle de Rasch, modèle linéaire mixte, données longitudinales, qualité de vie liée à la santé

Abstract :

In health sciences, it is frequent to deal with Patient Reported Outcomes (PRO) data, in particular for the evaluation of quality of life. Longitudinal data are often collected to allow analysing the evolution of an outcome over time. The most adequate strategy to analyse longitudinal latent variables, which can be either based on the calculated score or on Item Response Theory (IRT) models, remains to be identified. This strategy must take into account the latent characteristic of what PROs are intended to measure as well as the specificity of longitudinal designs with repeated measurements.

Our aim was to compare three methods to analyse such longitudinal latent variables through a simulation study. Two of the compared methods are based on the latent variable (IRT) whereas the last method is based on the score (classical approach).

In order to compare the methods, we evaluated the type I error, the power and the bias on the time effect estimation. These criteria can be affected by four parameters: (i) the sample size, (ii) the number of items, (iii) the correlation between two measures and (iv) the latent variable variance. Different scenarios were considered to evaluate the possible effect of these parameters.

This study shows different results in terms of type I error and power for the three methods. These criteria vary with the four identified parameters.

Keywords: Rasch model, linear mixed model, longitudinal data, health related quality of life

L'évaluation de mesures telles que la qualité de vie prend une place de plus en plus importante en recherche clinique et en épidémiologie, notamment en cancérologie. En effet, les études ne renseignent plus uniquement sur la santé du patient en termes de données cliniques et thérapeutiques mais permettent également de recueillir des données sur la santé perçue par le patient dont la qualité de vie est l'une des composantes majeures.

Les mesures subjectives, parfois appelées variables latentes sont évaluées à travers les réponses des patients à des items souvent regroupés en plusieurs dimensions au sein d'un questionnaire. Les méthodes couramment utilisées pour analyser ce type de données sont issues de deux approches : l'approche classique et celle issue de la Théorie de Réponse aux Items (IRT). L'approche classique est basée sur un score généralement calculé en sommant les réponses aux items. Il est supposé mesurer la vraie valeur du critère d'intérêt (qualité de vie). Dans la théorie de réponse aux items, la probabilité de répondre positivement à un item en fonction de la variable latente est utilisée pour estimer ce critère.

En recherche clinique et en épidémiologie, les données sont fréquemment recueillies de manière longitudinale afin de permettre de décrire l'évolution du critère d'intérêt dans le temps. Or, dans le cas de données longitudinales, la corrélation entre les mesures de chaque patient est à prendre en compte dans l'analyse à travers des méthodes appropriées.

A ce jour, la méthode la plus adéquate pour analyser des mesures subjectives recueillies longitudinalement reste à identifier. Elle peut être basée sur l'approche classique ou l'approche IRT et doit prendre en compte deux aspects fondamentaux : (i) le caractère latent du critère d'intérêt ainsi que (ii) la spécificité des données longitudinales.

Afin de répondre à cette problématique, une étude de simulation a été mise en place. Le but de cette étude était de comparer trois méthodes d'analyse de variables latentes longitudinales et d'évaluer leur intérêt respectif.

Nous nous sommes placés dans le cas d'items dichotomiques. Les données ont été simulées par un modèle de Rasch (Fischer et Molenaar (1995)), modèle le plus couramment utilisé en IRT. Pour se placer dans un cadre longitudinal, les individus ont été évalués à trois temps différents.

L'effet temps a été défini comme l'évolution standardisée des moyennes dans le temps, i.e.

effet temps = $\frac{\mu_{t+1} - \mu_t}{\sigma}$ avec μ_t la moyenne de la variable latente au temps t ($t=1,2$) et σ la variance de la variable latente. Sous H_0 , il n'y a pas d'effet temps. Sous H_1 , l'effet temps a été représenté par une évolution standardisée des moyennes de 0,2 entre deux visites consécutives.

Différents cas de figure ont été envisagés afin d'évaluer l'effet potentiel sur les performances des trois méthodes (i) de l'effectif de l'échantillon, (ii) du nombre d'items, (iii) de la corrélation entre les mesures et (iv) de la valeur de la variance de la variable latente. Des échantillons de 100 ou 200 individus ont été simulés. Le nombre d'items était de 4 ou 7. La structure de variance-covariance choisie pour la variable latente était de type autorégressive d'ordre 1. Avec cette structure, les variances à chaque temps sont considérées comme égales, les corrélations entre deux temps consécutifs le sont également ($\rho_{12} = \rho_{23} = \rho$). De plus, les réponses entre le premier et le troisième temps sont moins corrélées qu'entre deux temps consécutifs puisque $\rho_{13} = \rho^2$. La corrélation de la variable latente entre deux temps consécutifs pouvait avoir comme valeurs 0.4, 0.7 ou 0.9. Enfin, la

variance de la variable latente prenait comme valeurs 1, 4 ou 9. La combinaison des divers scénarios nous a amenés à considérer 36 cas différents. Pour chaque cas, 500 jeux de données simulées ont été étudiés.

Trois méthodes d'analyse des variables latentes longitudinales ont été évaluées. La première méthode (notée SM) est la plus usuelle. Elle est basée sur l'approche classique. Le score est calculé en sommant les réponses aux items. Puis, l'effet temps est estimé au moyen d'un modèle linéaire mixte expliquant le score (Verbeke et Molenberghs (2000)), modèle couramment employé pour analyser l'évolution d'une variable continue (score) au cours du temps. Les deux autres méthodes sont basées sur les modèles IRT. L'une (notée RM) consiste à estimer la valeur de la variable latente pour chaque individu par un modèle de Rasch puis à estimer l'effet temps par un modèle linéaire mixte expliquant cette variable latente. Avec l'autre méthode (notée LRM), l'effet temps est estimé avec un modèle de Rasch longitudinal (Embretson (1991)), modèle IRT adapté pour l'analyse de données longitudinales.

Afin de pouvoir comparer les différentes approches, trois critères ont été évalués pour chacune d'entre elles : le risque de première espèce, la puissance et le biais sur l'estimation de l'effet temps.

Les premiers résultats montrent que le risque de première espèce semble être maintenu proche de 5% pour la méthode SM ainsi que pour la méthode RM. La méthode LRM donne des résultats contradictoires. Le risque de première espèce est maintenu proche de 5% lorsque les variances sont peu élevées. Dans le cas où la variance de la variable latente a été fixée à 9, le risque de première espèce augmente fortement (jusqu'à 30%) avec l'augmentation du nombre d'items ou la diminution de la valeur de la corrélation.

Les méthodes SM et LRM donnent de bons résultats en terme de puissance. Lorsque le nombre d'individus et d'items est élevé, les puissances s'élèvent à plus de 80%. La méthode LRM semble, dans ce cas, mieux détecter la présence d'un effet temps. En effet, les puissances obtenues avec la méthode LRM sont légèrement plus élevées que celles obtenues avec la méthode SM. La puissance augmente avec le nombre d'items, le nombre d'individus, la valeur de la corrélation ou la valeur de la variance. Les puissances obtenues avec la méthode RM sont moins bonnes qu'avec les deux méthodes précédemment citées.

Bibliographie

- [1] Fischer G.H. et Molenaar I.W. (1995) Rasch models. Foundations, Recent Developments, and Applications, Springer-Verlag, New York.
- [2] Verbeke G. Et Molenberghs G. (2000) Linear Mixed Models for Longitudinal Data, Springer-Verlag, New York.
- [3] Embretson S.E. (1991) A multidimensional latent trait model for measuring learning and change, *Psychometrika*, 56, 495-515.